

Weekly Report

05/11/2015 - 05/17/2015

Jing XIA

May 17, 2015

1 Summary

This week I mainly focus on the DataScanner2 project (the second phase of data inspection).

2 Projects

2.1 Project 1 - DataScanner2

DataScanner2: A system that integrates various data quality metrics, covering all quality metrics categories and providing a quality-aware (clustering, correlation, outliers, complex patterns, image quality, feature preservation) visualization.

The scagnostics features are all implemented and integrated into the current DataScanner2 system. But because scagnostics was originally implemented for time series statistical data, it does not fit so well in our project: views on the same branch are not of similar shape.

We will keep the pixel-based and graph-based triangulation basis and develop new but features of sense to satisfy new needs.

2.2 Project 2 - NBA Game Visualization

-

2.3 Project 3 - FlowTracker

We start this project by reimplementing the “US Congress History” chart from XKCD (<http://xkcd.com/1127/>).

The basic data structure is a congress member class, which keeps records of all congressmen and their years of service in congress. We also made a new data structure with the extracted US congress dataset to meet the following functions:

1. There are six categories for each time step t . The data structure should satisfy quick extraction of size of all categories.
2. Of the six categories, each can be sourced from and evolving to 6 categories (other 5 categories and a non-congress category). The data structure should satisfy the tracing of all incoming and outgoing flows of all categories.

Guan is implementing this data structure and making basic visualization of it.

3 Paper Reading

-

4 To Do List

1. DataScanner2 — complete Scagnostics new features implementation.
2. DataScanner2 — define and implement quality metrics of single view.
3. FlowTracker — Basic visualization implementation.